the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2024) **120,** 9–18 doi: 10.1111/tpj.16989

Herbarium specimens as tools for exploring the evolution of fatty acid-derived natural products in plants

Emma Fitzgibbons[1], Jacob Lastovich[1], Samuel Scott[1], Nicole Groth[2], Amanda L. Grusz[2,3] and Lucas Busta[1,*]

[1]*Department of Chemistry and Biochemistry, University of Minnesota Duluth, Duluth, Minnesota, USA,*
[2]*Department of Biology, University of Minnesota Duluth, Duluth, Minnesota, USA, and*
[3]*Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA*

Received 8 May 2023; revised 20 May 2024; accepted 5 August 2024; published online 22 August 2024.
*For correspondence (e-mail bust0037@d.umn.edu).

SUMMARY

Plants synthesize natural products via lineage-specific offshoots of their core metabolic pathways, including fatty acid synthesis. Recent studies have shed light on new fatty acid-derived natural products and their biosynthetic pathways in disparate plant species. Inspired by this progress, we set out to develop tools for exploring the evolution of fatty-acid derived products. We sampled multiple species from all major clades of euphyllophytes, including ferns, gymnosperms, and angiosperms (monocots and eudicots), and we show that the compositional profiles (though not necessarily the total amounts) of fatty-acid derived surface waxes from preserved plant specimens are consistent with those obtained from freshly collected tissue in a semi-quantitative and sometimes quantitative manner. We then sampled herbarium specimens representing 57 monocot species to assess the phylogenetic distribution and evolution, of two fatty acid-derived natural products found in that clade: beta-diketones and alkylresorcinols. These chemical data, combined with analyses of 26 monocot genomes, revealed a co-occurrence (though not necessarily a causal relationship) between whole genome duplication and the evolution of diketone synthases from an ancestral alkylresorcinol synthase-like polyketide synthase. Limitations of using herbarium specimen wax profiles as proxies for those of fresh tissue seem likely to include effects from loss of epicuticular wax crystals, effects from preservation techniques, and variation in wax chemical profiles due to genotype or environment. Nevertheless, this work reinforces the widespread utility of herbarium specimens for studying leaf surface waxes (and possibly other chemical classes) and reveals some of the evolutionary history of fatty acid-derived natural products within monocots.

Keywords: plant chemistry, metabolic evolution, herbaria.

Linked article: This paper is the subject of a Research Highlight article. To view this Research Highlight article visit https://doi.org/10.1111/tpj.17038.

INTRODUCTION

Plant chemicals are a cornerstone of medicine and are also of great importance in agriculture and biotechnology (Wurtzel & Kutchan, 2016). Plant chemicals thus relate to several of the United Nations' Sustainable Development Goals, particularly those aimed at zero hunger as well as good health and well-being, since an advanced understanding of plant chemistry can help minimize the negative environmental impacts of agriculture and ensure food security in the face of climate change. Our knowledge of plant chemistry has advanced considerably in past decades and has recently been accelerated by advances in nucleotide sequencing technologies, leading to major advances in our understanding of specialized or lineage-specific metabolic pathways (Schenck & Busta, 2022). However, as studies of plant chemistry become wider in scope, they remain limited by a lack of access to tissue from diverse plant lineages, and researchers must spend considerable time, energy, and resources to obtain plant tissue for analyses. These difficulties are particularly pronounced when studying plants that are endangered, that grow in remote or difficult to access geographic areas, or that are challenging to cultivate. Until we identify straight-forward avenues by which to systematically access tissues from diverse plant species, our ability to advance fundamental knowledge of plant chemistry will be delayed. This, in turn, impedes the deployment of plant chemistry-based approaches to tackling climate change.

Outside of wild populations, much of global plant diversity is preserved within systematized collections,

including germplasm collections (seed banks), living collections (botanical gardens), and museum collections (herbaria). In recent decades, the plant voucher collections stored in herbaria have been transformed by digitization, allowing for systematic evaluation of collections spanning hundreds of millions of specimens collected around the globe over centuries (James et al., 2018; Soltis, 2017). Not only is this wealth of preserved plant specimens an incredible source of data for biodiversity, morphological, and phenological studies (Heberling & Burke, 2019; Johnson, Rebolleda-Gomez, & Ashman, 2019), it can also provide access to diverse taxa for chemical sampling of, for example, terpenoids, flavonoids, alkaloids, and glucosinolates (Cook et al., 2009; Eloff, 1999; Foutami, Mariager, Rinnan, Barnes, & Ronsted, 2018; Mithen, Bennett, & Marquez, 2010; Tasca et al., 2018). In the context of plant chemical research, we propose that using herbarium specimens could be especially useful for studies of metabolic evolution, for identifying lineages with variant metabolic machinery, as well as for discovering structural variants of known natural products and new natural products alike. Thus, herbaria have the potential to be a source of material for diverse studies that advance our understanding of plant chemistry.

The goal of this study was to evaluate the potential of herbarium specimens in studies of plant chemical diversity and plant metabolic evolution. To meet this goal, we chose to study fatty acid-derived natural products, a diverse group of compounds with both structural and bioactive properties (Scott, Cahoon, & Busta, 2022). These compounds can be found in cuticular waxes—mixtures of hydrophobic biochemicals that coat plant surfaces and can be remarkably resistant to degradation, as evidenced by their presence in fossils and geological samples (Damsté et al., 2003; Logan, Smiley, & Eglinton, 1995). We supposed that fatty acid-derived natural products could be extracted from the surfaces of small tissue fragments from herbarium specimens and then analyzed with gas chromatography–mass spectrometry as is currently performed with fresh tissue (Busta, Budke, & Jetter, 2016; Busta, Schmitz, Kosma, Schnable, & Cahoon, 2021; Guo, Li, Busta, & Jetter, 2018). In this report, we demonstrate that the wax chemical profiles obtained from preserved tissue (herbarium specimens) are semiquantitative, or in some cases quantitative, proxies for chemical profiles resulting from analyses of fresh tissue. Then, we used herbarium specimens to screen a wide range of monocot species for two classes of fatty acid-derived natural products, alkylresorcinols and beta-diketones, that function as defensive compounds (Cook et al., 2010; Ross et al., 2003), and enhance drought tolerance (Bi et al., 2017; Hen-Avivi et al., 2016; Zhang et al., 2015), respectively. Finally, we combined the chemical profiles from herbarium specimens with existing reference genome data for monocot species to provide a view into the evolutionary history of these two compounds and the importance of an alkylresorcinol synthase-like ancestor in that context. Based on this success, we suggest that phytochemists further develop partnerships with curators of natural history collections to advance large-scale studies of phytochemistry and to hasten the implementation of plant chemistry-based approaches to sustainable development goals.

RESULTS AND DISCUSSION

The objective of this study was to explore the use of herbarium specimens in studying the diversity and evolution of fatty acid-derived natural products in plant wax extracts. To meet this objective, three major activities were undertaken: (i) waxes from conspecific fresh and preserved specimens were compared across representatives of eudicot, monocot, gymnosperm, and fern lineages; (ii) preserved specimens were used to explore the distribution of alkylresorcinols and beta-diketones, two fatty acid-derived natural products, across the monocot lineage; and (iii) public sequence data were used to assess the evolution of alkylresorcinol and beta-diketone synthases across the monocot phylogeny.

Waxes from preserved specimens are highly similar to those from conspecific fresh tissue

An initial screen revealed striking similarities between the gas chromatography–mass spectrometry total ion chromatograms obtained from wax extracts from fresh and preserved tissues of *Turritis glabra* L. (Brassicaceae; Figure 1a). This suggested that information about compound occurrence, at least for some wax compounds, could potentially be obtained from analyses of preserved plant tissue. To evaluate this potential, our first objective was to perform a more thorough comparison of waxes from fresh and preserved specimens. We selected multiple species each from eudicot (3 species), monocot (4 species), gymnosperm (Soltis, 2017), and fern lineages (3; total 13 species; Figure 1b) and obtained fresh and preserved leaf tissue samples for three biologically independent individuals (i.e. three different plants, all three from the same populations for fresh tissues) from each of these species (total 72 tissue samples; where biologically independent samples were not available, distinct leaves from the same plant were used instead). Cuticular waxes were extracted from each tissue sample using chloroform and the extracts were analyzed with gas chromatography–mass spectrometry. Among the samples analyzed, we identified wax compounds belonging to eleven different wax chemical classes. Some of these chemical classes were ubiquitous classes that have been found, in some amount, on the surfaces of virtually all plant species, including very-long-chain *n*-alcohols, *n*-aldehydes, *n*-alkanes, alkyl esters, and fatty acids. We also identified wax compounds that are found only in certain lineages, including alkylresorcinols, as well as very-long-chain secondary alcohols, diols,

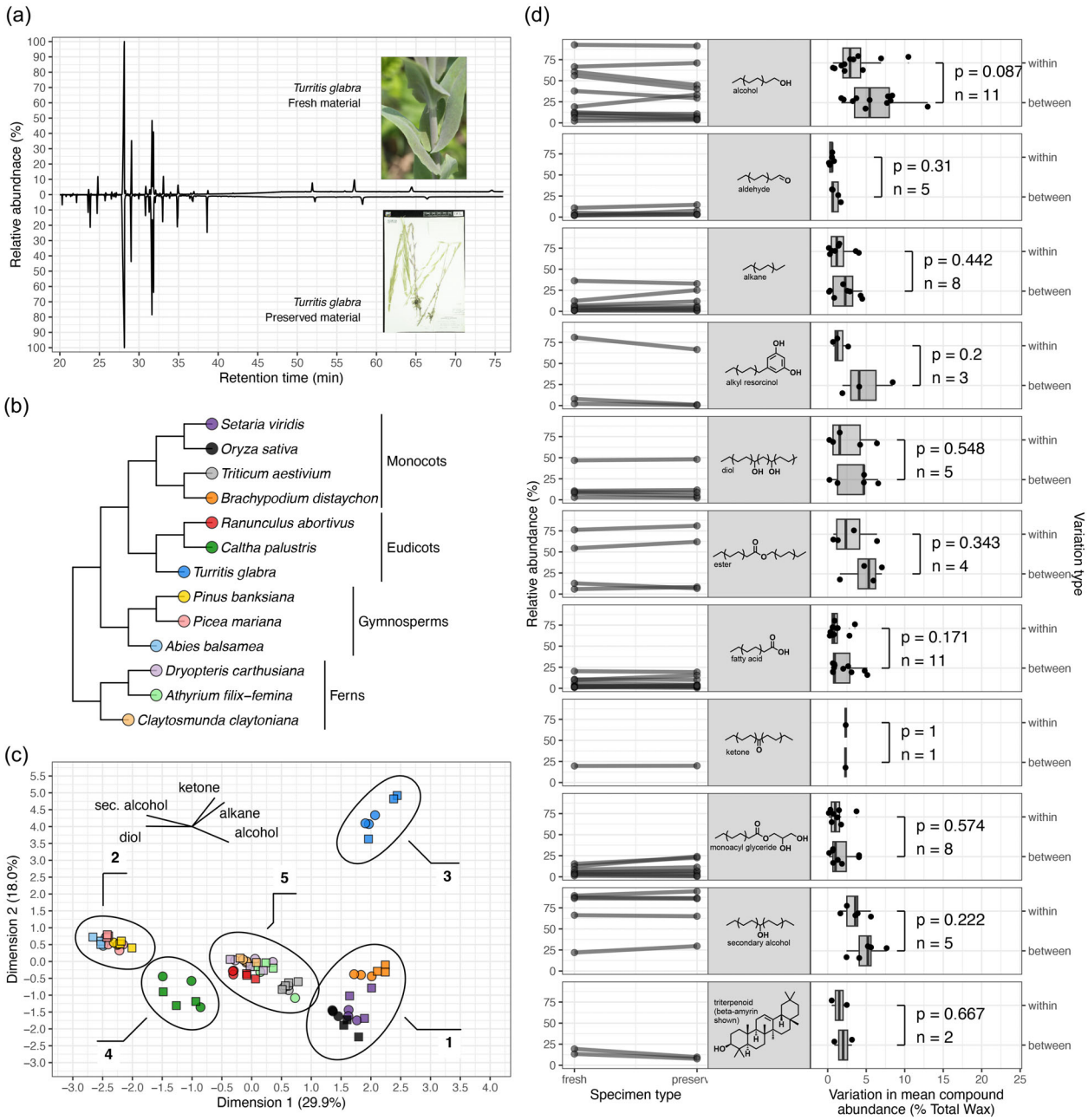

Figure 1. Comparison of waxes from fresh and preserved leaf tissue.
(a) Gas chromatography–mass spectrometry total ion chromatograms from analyses of cuticular waxes from fresh and preserved *Turritis glabra* tissue. The retention times of some late-eluting analytes do not exactly match because the samples were run under gas chromatography conditions with slightly different column lengths; however, the mass spectra for all corresponding compounds between the two samples were identical.
(b) Cladogram of the 12 species selected for analysis. This cladogram was derived by pruning a previously published megaphylogeny (Qian & Jin, 2016).
(c) Principal components analysis of the 72 wax samples analyzed. Circles correspond to fresh tissue, squares to preserved tissue, with colors that correspond to the species-based color scheme in b. Black circles indicate the major distinguishing chemotypes identified among the species analyzed. The line segments in the upper left-hand corner are an ordination plot illustrating the compound classes that are most strongly aligned with the principal components. Note that the clustering of Ranunculus arbortivus with the fern species in the center of the plot does not mean that their chemical profiles are necessarily similar, but rather that their chemical profiles are not distinguished by the variables comprising the first two principal components.
(d) Quantitative comparison of the relative abundance of wax compound classes on fresh and preserved plant material. Line diagrams on the left connect the relative abundance of each compound class in conspecific fresh and preserved tissues. Slopes of the lines indicate the degree of difference in the relative abundance of a particular compound class between conspecific fresh and preserved specimens. The number of lines shown for each compound class indicates the number of species in which that compound class was found. The panels in the middle show the general structure of each compound class. Box plots on the right show the amount of variation in mean compound class relative abundance within and between a set of independent samples.

ketones, monoacyl glycerides, and triterpenoids. Overall, the data suggest that preserved plant specimens could be a source of diverse wax compound classes. Previous comparisons of volatile terpenoids from herbarium specimens and fresh tissues revealed that herbarium specimens had a similar distribution of terpenoid compounds (terpenoid "composition"), but not necessarily similar total terpenoid amounts (Foutami et al., 2018). It seems likely that surface waxes would behave similarly, given that some species waxes accumulate in thick surface layers (epicuticular wax crystals) that can be physically removed with even a gentle touch, let alone actions associated with collection, pressing, mounting, and storing. For that reason, the remainder of the analyses presented here focus on wax composition, not total wax amounts.

To evaluate the potential of waxes from preserved plant specimens to serve as quantitative proxies for extracts from freshly collected tissue, quantitative chemical profiles from the fresh and preserved tissue of 12 representative plant species were compared (Table S1). We conducted a principal components analysis of the chemical profiles and found that the 72 samples fell into five general clusters or chemotypes (Figure 1c, numbered ovals). Four of the clusters were distinguished from each other by being enriched in one or more compound classes, with cluster 1 being *n*-alcohol rich, clusters 2 and 3 being enriched in secondary alcohols and diols, and cluster 4 being enriched in alkanes and ketones (Figure 1c), compare cluster position with inset ordination plot. The central cluster (cluster 5) was not distinguished by being particularly rich in any compound class relative to the other clusters. Importantly, samples from fresh and preserved tissue of the same species grouped tightly together in the principal components analysis. This clustering suggested that, even though some of the preserved specimens were nearly 100 years old (Table S2), the chemistries of fresh and preserved specimens from the same species are generally more similar to one another than to the chemistry of other species (Figure 1c).

Since different wax compound classes can degrade at different rates (i.e., different compound classes might be preserved to different degrees), the present analysis was conducted on a compound class-by-compound class basis for each species. To visualize such trends in preservation, a plot was created in which the mean relative abundance of each compound class in fresh tissue was connected with a line to the mean relative abundance of that class in preserved tissue from that species (Figure 1d). The lines in the plot were relatively flat, indicating that the relative abundance of each compound class detected was highly similar in extracts of conspecific fresh and preserved specimens and suggested that a statistical analysis should be devised to test for differences between these relative abundance values. However, it is well established that cuticular wax chemistry can change quantitatively in response to environmental stimuli (Kosma et al., 2009; Lewandowska, Keyl, & Feussner, 2020; Shepherd, 2006), and it seemed highly likely that such variation might preclude direct quantitative comparisons of preserved and fresh tissue chemistry. Accordingly, to determine whether wax compounds differed in their quantitative contribution to overall wax mixtures, it was not the mean abundances between fresh and preserved tissue that were compared, but rather the variance in the abundance of each compound class between replicates of a given tissue type (fresh vs. preserved) versus the variance between fresh and preserved tissue. Under this scheme, if a compound class regularly degraded in a quantifiable way, the variance between fresh and preserved would be significantly greater than the variance within a tissue type. For the compound classes tested here, we compared these variances and found no significant difference for any compound class (Wilcox tests, $\alpha = 0.05$, Figure 1d). This absence of differences suggested that each of these compound classes could, at least in the cases tested here, indeed be preserved at a quantitative level.

Although our analyses suggested semi-quantitative, or in some cases even quantitative similarity between the wax composition (though not necessarily coverage) of fresh and preserved wax specimens, we would like to emphasize several important limitations of our method. First, more wax classes have been reported than those we tested here, and semi-quantitative preservation may not be the case for all species or compound classes. In addition, epicuticular and intracuticular wax compositions differ in some species and thus the inadvertent removal of epicuticular wax enriched in a given compound class would almost certainly distort correspondence between fresh and preserved tissue. We also did not test for any effects that preservation techniques may have on wax profiles (such as the use of ethanol in preservation), and cannot rule out that the wax chemical profiles of preserved specimens would be affected by specific preservation techniques. Finally, there can be considerable variation in wax composition between members of the same species due to either genotype or environmental effects. Such differences could also distort correspondence between fresh and preserved tissues in some cases. One of the effects of these limitations could be difficulty in the ability to distinguish among wax profiles from species of the same genus, though such would depend on the amount of variation in the wax compositions of the species in question. However, even considering these limitations, the data presented here suggest that herbarium specimens are a suitable first approximation for the quantitative wax compositions found on fresh plant leaf tissue. It seems likely that this approach will also work for other classes of surface lipids, notably cutin, a polymer whose diversity is little studied compared with waxes.

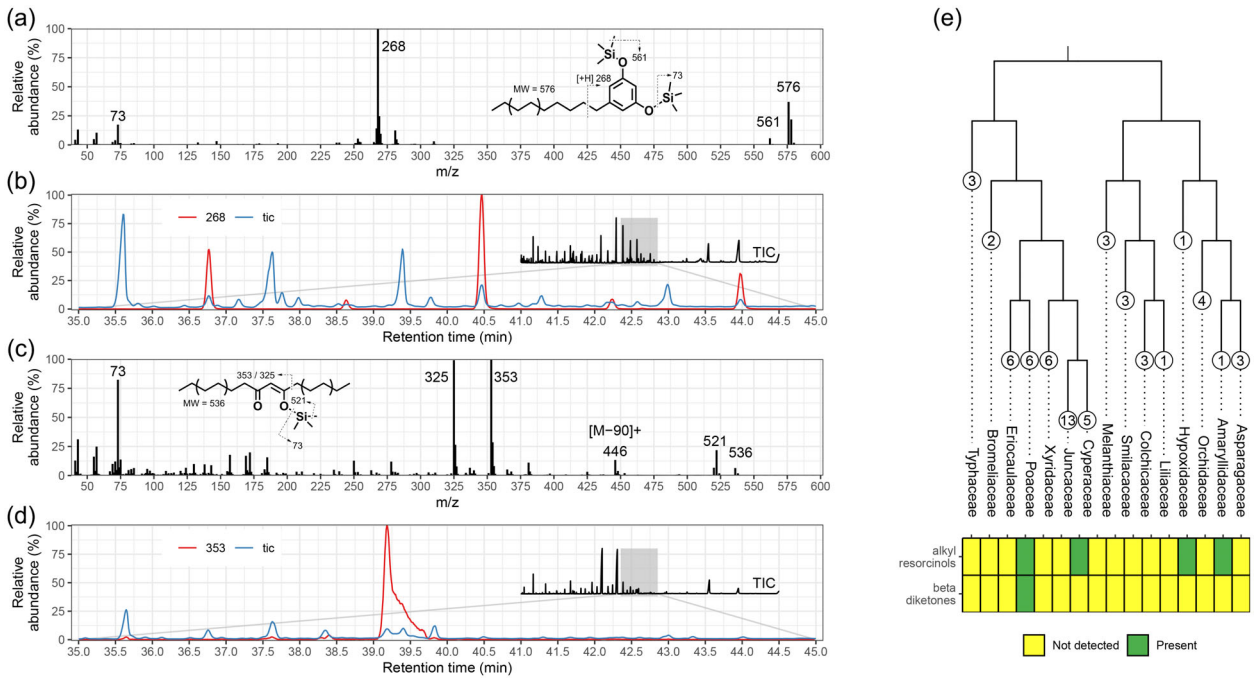

Figure 2. Occurrence of alkylresorcinol and beta-diketone fatty acid natural products in monocot lineages.
(a) Mass spectrum of an alkylresorcinol, with a prominent mass peak at *m/z* 268.
(b) Total and extracted (*m/z* 268) ion chromatograms of wax extracts from preserved tissues of *Trichophorum cespitosum*. The full total ion chromatogram is inset.
(c) Mass spectrum of a beta-diketone, with prominent mass peaks at *m/z* 325 and *m/z* 353.
(d) Total and extracted (*m/z* 353) ion chromatograms of wax extracts from preserved tissues of *Hordeum jubatum*. The full total ion chromatogram is inset.
(e) Cladogram of fifteen monocot families from which wax extracts from preserved tissues were analyzed. The numbers in the circles above each family indicate the number of species in that family that were analyzed. The heat map below the family names indicates the families in which alkylresorcinols and beta-diketones were found, with green squares indicating that the compound class was found in that lineages and yellow indicating that the compound was not detected in that lineage. The data acquired for each species in each of these lineages is like that shown in panels a–d and is presented in Figure S2.

Based on preserved specimen chemistry, alkylresorcinols are more widely distributed in monocots than are beta-diketones

Having identified that wax extracts from preserved specimens are reasonable proxies for extracts from conspecific fresh tissue, our next objective was to use herbarium specimen wax extracts to study the evolution of fatty acid-derived natural products. In particular, we targeted alkylresorcinols and beta-diketones, two compound classes that have been reported from monocot lineages (Mikkelsen, 1979; Racovita & Jetter, 2016; Sun et al., 2020), but whose evolution has not been studied in detail. Alkylresorcinols are defensive antifungal and antimicrobial compounds with potential anticancer effects (Cook et al., 2010; Ross et al., 2003), while beta-diketones enhance leaf survival and reduce water loss under drought stress (Bi et al., 2017; Hen-Avivi et al., 2016; Zhang et al., 2015). We began by examining wax extracts from preserved tissues of species previously reported to produce alkylresorcinols or beta-diketones. Alkylresorcinols, as trimethylsilyl derivatives, have a prominent peak in their mass spectra at *m/z* 268 (Figure 2a), so single ion chromatograms were extracted for that mass-to-charge ratio to search for alkylresorcinols. In a wax extract from preserved tissues of *Trichophorum cespitosum* (L.) Hartm., which was previously reported to contain alkylresorcinols (Avesjs et al., 2002), alkylresorcinols were indeed identified as a homologous series (Figure 2b).

Beta-diketones, as trimethylsilyl derivatives, have prominent ions in their mass spectra corresponding to alpha-cleavages adjacent to their secondary functional groups (Figure 2c). Accordingly, to search for diketones, gas chromatography runs were conducted in scan mode and then ion chromatograms were extracted for the ions that would be generated by a variety of possible beta-diketone structures including most common beta-diketone structures from grasses (Figure S1). For example, hentriacontan-14,16-dione was identified by its prominent mass peaks at *m/z* 325 and *m/z* 353 in preserved tissues of *Hordeum jubatum* L., a species previously reported to contain diketones (Baum, Tulloch, & Bailey, 1989). Thus, both alkylresorcinols and beta-diketones can be detected in wax extracts from preserved plant tissues,

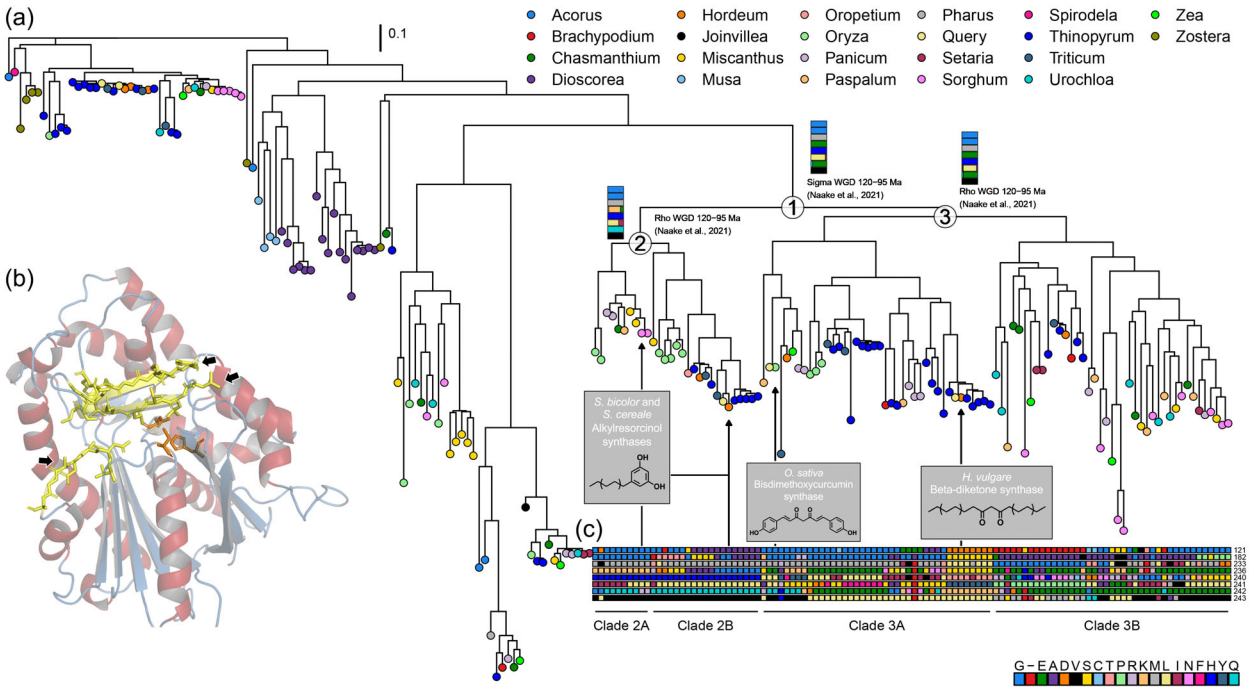

Figure 3. Characteristics of extant and ancestral polyketide synthases.
(a) Maximum likelihood tree of type-III polyketide synthase amino acid sequences from 26 monocot genomes. Tip color indicates the family in which the gene was found, as shown in the legend. The nodes labeled with numbered circles "1", "2", and "3" indicate whole genome duplications, with whole genome duplication at "1" preceding the divergence of clades 2 and 3 and whole genome duplications at "2" and "3" giving rise to 2A/2B and 3A/3B.
(b) A computer-generated model of a Poaceae type-III polyketide synthase based on structurally similar crystal structures. The three orange residues are catalytic residues, and the eight yellow residues indicate amino acids near the active site pocket.
(c) Colored squares indicating the identity of the eight active site-adjacent amino acids (shown in yellow in b) in each of the proteins descended from the node labeled "1" in the phylogeny. Numbers on the right-hand side of the colored squares indicate the position in our alignment containing each of these residues. The positions of those residues in the structural model in b are indicated with yellow in the model as denoted with black arrows. The colors correspond to a particular amino acid as indicated in the legend in the lower right-hand corner of the figure. The horizontal bar plots at each whole genome duplication node in the phylogeny indicate the probability of the eight amino acid residues near the active site pocket in the ancestral sequences predicted for that node.

and this process is facilitated with the use of extracted ion chromatograms.

Having confirmed that alkylresorcinols and beta-diketones can be detected in wax extracts from preserved plant tissues, our next objective was to determine their relative times of evolution by determining whether alkylresorcinols or beta-diketones are more widely distributed across monocots. To meet that goal, wax extracts were prepared from preserved tissue of fifty seven monocot herbarium specimens spanning fifteen monocot families (Figure 2e; Table S3). Each extract was analyzed with gas chromatography–mass spectrometry to generate data similar to that shown in Figure 2a–d. Across these samples, alkylresorcinols were detected in at least one species from each of the families Poaceae (3 out of 6 species sampled), Cyperaceae (5/5), Orchidaceae (1/4), and Asparagaceae (2/3). In contrast, beta-diketones were only found on species within the Poaceae (1 out of 6 species). This suggests that alkylresorcinols are more widespread in monocots than are beta-diketones and that the ability to synthesize alkylresorcinols evolved before the ability to synthesize beta-diketones in this lineage.

A phylogeny of monocot polyketide synthases suggests genome duplication co-occurred with diversification of ancestral alkylresorcinol synthase-like genes and the evolution of diketone synthesis

So far, the herbarium specimen-derived chemical data suggested that alkylresorcinol synthesis predated beta-diketone biosynthesis in monocots. In addition, previous studies had identified alkylresorcinol and beta-diketone synthesis genes in monocots (Cook et al., 2010; Sun, 2023; Sun et al., 2020), all of which are type-III polyketide synthases. Accordingly, our next objective was to use these characterized sequences and the previously assembled reference genomes of 26 monocot species (Table S4) to place our chemical data into a genomic context and to further explore the evolutionary history of beta-diketones. First, we used the amino acid sequences of the characterized alkylresorcinol and beta-diketone synthases as queries for a BLAST search and obtained a list of type-III polyketide synthases from the 26 monocot genomes. We filtered out low-quality hits as well as hits that were clearly annotated as something other than a

polyketide synthase, for example, 3-ketoacyl CoA synthases. The sequences of the remaining hits (Table S5) were aligned, and the alignment was trimmed to remove positions with a high proportion of gaps (more than 70% gaps) or positions that had very low conservation (less than 30% conserved). From the trimmed alignment, we inferred a maximum likelihood phylogeny (Figure 3a; Table S6). The overall topology of our phylogeny was highly similar to the monocot region of a previously reported type-III polyketide synthase phylogeny (Naake, Maeda, Proost, Tohge, & Fernie, 2021). In our phylogeny, the previously characterized monocot alkylresorcinol and beta-diketone synthases were located in four large clades made up of sequences from the Poaceae (Clades 2A/2B and 3A/3B, Figure 3). These clades diversified after previously reported whole genome duplications in this lineage (nodes labeled "1", "2", and "3"; Figure 3; (Baum et al., 1989)). Gene duplications, from the single gene to whole genome level, have previously been associated with the functional diversification of genes involved in lineage-specific metabolism (Bai et al., 2022; Feng et al., 2019; Schenck & Busta, 2022; Yonekura-Sakakibara, Higashi, & Nakabayashi, 2019). This precedent, along with the characterized alkylresorcinol and beta-diketone synthases arising after the whole genome duplications, suggests that whole genome duplication could have played a role in the functional diversification of type-III polyketide synthases in monocots. To test this idea, we used a subset of the 26 monocot genomes to examine patterns in synteny across the genomes of these monocots. However, our analyses did not detect any strong evidence for whole genome duplication as a cause of polyketide synthase gene family expansion. Still, the temporal co-occurrence of the duplications and gene family expansion is intriguing.

The finding that characterized alkylresorcinol synthases and a characterized beta-diketone synthase arose from a common polyketide synthase ancestor shortly after whole genome duplication raised questions about the catalytic abilities of the common ancestor. To learn more about this ancestor, as well as other ancestral polyketide synthase sequences, we first built a structural model of a plant type-III polyketide synthase and used it to identify positions in our type-III polyketide synthase amino acid alignment that correspond to active site-adjacent residues (Figure 3b). We then mapped the identity of those active site-adjacent residues onto the phylogeny (Figure 3c) and examined how those identities were distributed across the four clades arising after the two whole genome duplications. The two characterized alkylresorcinol synthases and other members of both clades 2A and 2B had characteristic residues in certain position of the alignment (M233, H240, Q242, and V243; Figure 3c). In contrast, the characterized beta-diketone synthase did not have any of these characteristic active site residues (it instead had L233, N240, K242, and L243), something it shared in common with its closest neighbors in our phylogeny—genes from *Thinopyrum intermedium* (Host) Barkworth & D.R. Dewey, suggesting that species as a beta-diketone producer and one or more of those genes as beta-diketone synthases. Thus, analysis of alkylresorcinol and beta-diketone synthase amino acid sequences revealed that the identities of the amino acids surrounding their active sites have diverged considerably.

Since eight active site-adjacent residues (yellow residues indicated with black arrows in Figure 3b, also represented as colored squares in Figure 3c) so clearly distinguished alkylresorcinol and beta-diketone synthases, we supposed that it might be possible to distinguish alkylresorcinol synthase-like ancestral sequences from beta-diketone synthase-like sequences. Accordingly, we used maximum likelihood ancestral state reconstruction to estimate the identity of the same eight active site-adjacent residues at the three nodes of the phylogeny that corresponded to whole genome duplications (Table S7). At each of these nodes, the ancestral state reconstruction algorithm was able to estimate the identity of the eight residues with high certainty, and each predicted ancestor had characteristics that were strikingly similar to characterized alkylresorcinol synthases (M233, H240, Q242, and V243) and absent in the beta-diketone synthase (Figure 3c, compare with colored bar charts at duplications). This finding strongly suggests that the ancestral polyketide synthase that gave rise to the sequences in clades 2A, 2B, 3A, and 3B (that is, both alkylresorcinol synthases and the beta-diketone synthase) was very alkylresorcinol synthase-like, if not an outright alkylresorcinol synthase.

CONCLUSION

Here, we demonstrate that cuticular waxes from herbarium specimens can be sources of fatty acid-derived natural products in proportions (i.e. with compositions) that are semiquantitatively consistent with the chemistry of fresh tissue. We then surveyed monocot wax chemicals and demonstrated that alkylresorcinols are far more widespread across the monocot phylogeny than beta-diketones, which were only recovered from within Poaceae. We used ancestral state reconstruction and genomic analyses of monocot type-III polyketide synthases to reveal that, within the Poaceae, an ancestral alkylresorcinol synthase-like gene appears to have diversified shortly after whole genome duplication, and that some products of that diversification beta-diketone synthases. Our analyses also reveal several groups of polyketide synthases with virtually unexplored functions. It will be interesting to see what products' synthesis they may encode. Overall, our findings suggest that herbarium specimens are a useful source of plant waxes for analytical study, though limitations are imposed by several factors, including effects from loss of epicuticular wax crystals, effects from preservation

techniques, and variation in wax chemical profiles due to genotype or environment. Still, this method has the potential to provide access to otherwise inaccessible dimensions of natural variation (Berenbaum & Zangerl, 1998; Lang, Willems, Scheepens, Burbano, & Bossdorf, 2019; Zangerl & Berenbaum, 2005), including analyses of waxes from plant tissues that were collected in the same location but separated by decades or even centuries of time.

EXPERIMENTAL PROCEDURES

Plant material and chemical analysis

Fresh plant material was collected from local green areas. Preserved plant samples were provided on loan by the Olga Lakela Herbarium (DUL) at the University of Minnesota-Duluth. To prepare a wax sample, whether from fresh or preserved tissue, a large hole punch was used to collect a disc of leaf tissue (2 cm in diameter) that was stored in a sealed scintillation vial. The discs of leaf tissue were spiked with 10 μg of tetracosane internal standard then extracted twice with 2 mL chloroform. The extracts were pooled, and the chloroform was allowed to evaporate. The residue was then transferred to a GC vial insert, derivatized in 50 μL of 1:1 N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) and pyridine, and incubated (70°C for 45 min). Samples were analyzed on a 7890B Network GC (Agilent) equipped with an 7693A Autosampler (Agilent) equipped with a split/splitless injector and an HP-5 capillary column (Agilent, length 30 m × 0.250 mm × 0.25 μm film thickness); 2 μL of sample was injected on-column with He as the carrier gas with a flow rate of 1 mL/min. The initial temperature of the GC oven was 50°C and held for 2 min, followed by the first ramp at a rate of 40°C/min until it reached 200°C and was held for 2 min, then ramped at a rate of 3°C/min until 320°C and held for 30 min. The total run time for each sample was 77.75 min. with a solvent delay of 8 min. The analytes were detected using an Agilent 5977B (GC/MSD) Mass Selective Detector (EI 70 eV; m/z 40–800, 2 scans/s). Peaks were identified by comparing mass spectra again those of authentic standards, previously published spectra for authentic standards, or commercial mass spectral libraries.

Bioinformatics and data analysis

Species and family trees (Figures 1b and 2e) were obtained by sampling from a published megaphylogeny (Qian & Jin, 2016). Principal components analysis was performed using the R package FactoMineR (Lê, Josse, & Husson, 2008). BLAST searches, alignments, trimming, tree construction and visualization, and ancestral state reconstruction were performed as previously described (Damsté et al., 2003) using the R packages "msa," "gBlocks," "phangorn," and "ggtree" (Bodenhofer, Bonatesta, Horejš-Kainrath, & Hochreiter, 2015; Schliep, 2011; Talavera & Castresana, 2007; Yu, Smith, Zhu, Guan, & Lam, 2017). Parameters for filtering BLAST hits were length aligned with query >250; hit length >300 and <600, e_value <1E-50. Statistical tests were performed using the R package "rstatix." Structural modeling was performed as previously described (Busta et al., 2018) using Phyre2 (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015) in intensive mode. The base model selected by Phyre2 for the Hordeum vulgare Type III polyketide synthase (beta-diketone synthase) structural model presented here was an *Oryza sativa* Type III PKS (Morita et al., 2010), PDB entry 3ALE. The model was visualized using PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC) and positions in the Type III PKS alignment that corresponded to active site residues were identified by comparing the alignment against the sequence of the model shown in PyMOL. Synteny analyses were performed using GENESPACE (Lovell et al., 2022) using genomes from Phytozome for the 26 monocot species, provided less than 5% of their genes were on small scaffolds, leaving 20 genomes remaining for synteny analysis.

ACKNOWLEDGEMENTS

The authors wish to acknowledge support in the form of startup funds to LB from the Swenson College of Science and Engineering. We also extend our thanks to the University of Minnesota Duluth for supporting JL through the Undergraduate Research Opportunities Program. EF is grateful for the support of the Department of Chemistry and Biochemistry and the UMD Chemistry Master's Program. ALG is also grateful to NSF-DEB 2232106 for generous financial support. Specimens analyzed in this study were generously supplied by the Olga Lakela Herbarium (DUL). We collectively acknowledge that the University of Minnesota Duluth is located on the traditional, ancestral, and contemporary lands of Indigenous people. The University resides on land that was cared for and called home by the Ojibwe people, before them the Dakota and Northern Cheyenne people, and other Native peoples from time immemorial. Ceded by the Ojibwe in an 1854 treaty, this land holds great historical, spiritual, and personal significance for its original stewards, the Native nations, and peoples of this region. We recognize and continually support and advocate for the sovereignty of the Native nations in this territory and beyond. By offering this land acknowledgment, we affirm tribal sovereignty and will work to hold the University of Minnesota Duluth accountable to American Indian peoples and nations.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All relevant data can be found within the manuscript and its supporting materials.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Fragmentation patterns of trimethylsilyl-derivatized beta-diketones. Each row shows structures for possible beta-diketones isomers and major fragments they may generate during mass spectrometry.

Figure S2. Single-ion chromatograms of alkylresorcinols in diverse species. Each panel shows the single-ion chromatogram (m/z 268, characteristic for alkylresorcinols) traces for wax sample chromatograms of the species indicated on the right.

Table S1. Wax chemical data for the thirteen species used in the investigation of fresh and preserved tissue wax chemistry. Listed is the abundance (percent total wax) of each compound found in each species for the twelve species used in the investigation of fresh and preserved tissue wax chemistry.

Table S2. Herbarium specimens used to assess wax chemical preservation. Listed are the family, taxon, genus, species, replicate, and collector number. Entries marked with a dagger indicate instances in which biologically independent samples were not

available and so independent leaves from the same individual were used instead.

Table S3. List of Poaceae herbarium specimens surveyed for beta-diketones and alkylresorcinols. Listed are the family, taxon, accession number for each specimen that was examined. The fourth and fifth columns indicate whether beta-diketones or alkylresorcinols were found in that lineage.

Table S4. Public sequence data used in this study. Listed are the genus and species names of taxa whose genomes were used in this study. The data were obtained from Phytozome13 (https://phytozome-next.jgi.doe.gov/), with the file names listed alongside species names.

Table S5. Metadata for Poales type III PKS sequences. Included are the accession number for each sequence, the genus and species name for the genome from which it was obtained, the length of the sequence, as well as the BLAST parameters obtained when it was found using a BLAST search: length_aligned_with_query, percent_identity, e_value, and bitscore.

Table S6. Poaceae Type III Polyketide Synthase Phylogeny. A fortified tree of the phylogeny shown in Figure 3. This is a maximum likelihood phylogeny built from the alignment in Data S1. Note that the R function "as.phylo()" can be used to convert this table into a "phylo" object which can then be saved as a newick or nexus file.

Table S7. Ancestral States of Type III Polyketide Synthases from Poaceae. The probability of the states of each position in the alignment in Data S1 for each of the nodes of the phylogenetic tree in Figure 3, which is included in Table S6.

Data S1. Alignment of type III PKSs from Poaceae. A fasta file containing the alignment built from the sequences described in Table S5.

REFERENCES

Avesjs, L.A. *et al.* (2002) 5-n-alkylresorcinols as biomarkers of sedges in an ombrotrophic peat section. *Organic Geochemistry*, **33**, 861–867.

Bai, G., Chen, C., Zhao, C., Zhou, T., Li, D., Zhou, T. *et al.* (2022) The chromosome-level genome for *Toxicodendron vernicifluum* provides crucial insights into anacardiaceae evolution and urushiol biosynthesis. *iScience*, **25**, 104512.

Baum, B.R., Tulloch, A.P. & Bailey, G.L. (1989) Epicuticular waxes of the genus hordeum: a survey of their chemical composition and ultrastructure. *Canadian Journal of Botany*, **67**, 3219–3226.

Berenbaum, M.R. & Zangerl, A.R. (1998) Chemical phenotype matching between a plant and its insect herbivore. *Proceedings of the National Academy of Sciences*, **95**, 13743–13748.

Bi, H., Kovalchuk, N., Langridge, P., Tricker, P.J., Lopato, S. & Borisjuk, N. (2017) The impact of drought on wheat leaf cuticle properties. *BMC Plant Biology*, **17**, 85.

Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C. & Hochreiter, S. (2015) Msa: an r package for multiple sequence alignment. *Bioinformatics*, **31**, 3997–3999.

Busta, L., Budke, J.M. & Jetter, R. (2016) The moss *Funaria hygrometrica* has cuticular wax similar to vascular plants, with distinct composition on leafy gametophyte, calyptra and sporophyte capsule surfaces. *Annals of Botany*, **118**, 511–522.

Busta, L., Schmitz, E., Kosma, D.K., Schnable, J.C. & Cahoon, E.B. (2021) A co-opted steroid synthesis gene, maintained in sorghum but not maize, is associated with a divergence in leaf wax chemistry. *Proceedings of the National Academy of Sciences*, **118**, e2022982118.

Busta, L. *et al.* (2018) Identification of genes encoding enzymes catalyzing the early steps of carrot polyacetylene biosynthesis. *Plant Physiology*, **178**, 1507–1521.

Cook, D., Gardner, D.R., Pfister, J.A., Welch, K.D., Green, B.T. & Lee, S.T. (2009) The biogeographical distribution of duncecap larkspur (*Delphinium occidentale*) chemotypes and their potential toxicity. *Journal of Chemical Ecology*, **35**, 643–652.

Cook, D., Rimando, A.M., Clemente, T.E., Schröder, J., Dayan, F.E., Nanayakkara, N.P.D. *et al.* (2010) Alkylresorcinol synthases expressed in sorghum bicolor root hairs play an essential role in the biosynthesis of the allelopathic benzoquinone sorgoleone. *The Plant Cell*, **22**, 867–887.

Damsté, J.S., Rampen, S., Irene, W., Rijpstra, C., Abbas, B., Muyzer, G. *et al.* (2003) A diatomaceous origin for long-chain diols and mid-chain hydroxy methyl alkanoates widely occurring in quaternary marine sediments: indicators for high-nutrient conditions. *Geochimica et Cosmochimica Acta*, **67**, 1339–1348.

Eloff, J. (1999) Is it possible to use herbarium specimens to screen for antibacterial components in some plants. *Journal of Ethnopharmacology*, **67**, 355–360.

Feng, T. *et al.* (2019) FAD2 gene radiation and positive selection contributed to polyacetylene metabolism evolution in campanulids. *Plant Physiology*, **181**, 714–728.

Foutami, I.J., Mariager, T., Rinnan, R., Barnes, C.J. & Ronsted, N. (2018) Hundred fifty years of herbarium collections provide a reliable resource of volatile terpenoid profiles showing strong species effect in four medicinal species of salvia across the mediterranean. *Frontiers in Plant Science*, **9**, 1877.

Guo, Y., Li, J.J., Busta, L. & Jetter, R. (2018) Coverage and composition of cuticular waxes on the fronds of the temperate ferns *Pteridium aquilinum*, *Cryptogramma crispa*, *Polypodium glycyrrhiza*, *Polystichum munitum* and *Gymnocarpium dryopteris*. *Annals of Botany*, **122**, 555–568.

Heberling, J.M. & Burke, D.J. (2019) Utilizing herbarium specimens to quantify historical mycorrhizal communities. *Applications in Plant Sciences*, **7**, e01223.

Hen-Avivi, S., Savin, O., Racovita, R.C., Lee, W.S., Adamski, N.M., Malitsky, S. *et al.* (2016) A metabolic gene cluster in the wheat w1 and the barley cer-cqu loci determines beta-diketone biosynthesis and glaucousness. *The Plant Cell*, **28**, 1440–1460.

James, S.A., Soltis, P.S., Belbin, L., Chapman, A.D., Nelson, G., Paul, D.L. *et al.* (2018) Herbarium data: global biodiversity and societal botanical needs for novel research. *Applications in Plant Sciences*, **6**, e1024.

Johnson, A.L., Rebolleda-Gomez, M. & Ashman, T.L. (2019) Pollen on stigmas of herbarium specimens: a window into the impacts of a century of environmental disturbance on pollen transfer. *The American Naturalist*, **194**, 405–413.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.E. (2015) The phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, **10**, 845–858.

Kosma, D.K., Bourdenx, B., Bernard, Á., Parsons, E.P., Lü, S., Joubès, J. *et al.* (2009) The impact of water deficiency on leaf cuticle lipids of arabidopsis. *Plant Physiology*, **151**, 1918–1929.

Lang, P.L.M., Willems, F.M., Scheepens, J.F., Burbano, H.A. & Bossdorf, O. (2019) Using herbaria to study global environmental change. *The New Phytologist*, **221**, 110–122.

Lê, S., Josse, J. & Husson, F. (2008) Factominer: an r package for multivariate analysis. *Journal of Statistical Software*, **25**, 1–18.

Lewandowska, M., Keyl, A. & Feussner, I. (2020) Wax biosynthesis in response to danger: its regulation upon abiotic and biotic stress. *The New Phytologist*, **227**, 698–713.

Logan, G., Smiley, C.J. & Eglinton, G. (1995) Preservation of fossil leaf waxes in association with their source tissues, clarkia, northern Idaho, USA. *Geochimica et Cosmochimica Acta*, **59**, 751–763.

Lovell, J.T., Sreedasyam, A., Schranz, M.E., Wilson, M., Carlson, J.W., Harkess, A. *et al.* (2022) GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*, **11**, e78526.

Mikkelsen, J.D. (1979) Structure and biosynthesis of beta-diketones in barley spike epicuticular wax. *Carlsberg Research Communications*, **44**, 133–147.

Mithen, R., Bennett, R. & Marquez, J. (2010) Glucosinolate biochemical diversity and innovation in the brassicales. *Phytochemistry*, **71**, 2074–2086.

Morita, H., Wanibuchi, K., Nii, H., Kato, R., Sugio, S. & Abe, I. (2010) Structural basis for the one-pot formation of the diarylheptanoid scaffold by curcuminoid synthase from *Oryza sativa*. *Proceedings of the National Academy of Sciences*, **107**, 19778–19783.

Naake, T., Maeda, H.A., Proost, S., Tohge, T. & Fernie, A.R. (2021) Kingdom-wide analysis of the evolution of the plant type III polyketide synthase superfamily. *Plant Physiology*, **185**, 857–875.

Qian, H. & **Jin, Y.** (2016) An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *Journal of Plant Ecology*, **9**, 233–239.

Racovita, R.C. & **Jetter, R.** (2016) Identification of polyketides in the cuticular waxes of *Triticum aestivum* cv. Bethlehem. *Lipids*, **51**, 1407–1420.

Ross, A.B., **Shepherd, M.J.**, **Schüpphaus, M.**, **Sinclair, V.**, **Alfaro, B.**, **Kamal-Eldin, A.** ***et al.*** (2003) Alkylresorcinols in cereals and cereal products. *Journal of Agricultural and Food Chemistry*, **41**, 4111–4118.

Schenck, C.A. & **Busta, L.** (2022) Using interdisciplinary, phylogeny-guided approaches to understand the evolution of plant metabolism. *Plant Molecular Biology*, **109**, 355–367.

Schliep, K.P. (2011) Phangorn: phylogenetic analysis in r. *Bioinformatics*, **27**, 592–593.

Scott, S., **Cahoon, E.B.** & **Busta, L.** (2022) Variation on a theme: the structures and biosynthesis of specialized fatty acid natural products in plants. *The Plant Journal*, **111**, 954–965.

Shepherd, T. (2006) D Wynne Griffiths, the effects of stress on plant cuticular waxes. *The New Phytologist*, **171**, 469–499.

Soltis, P.S. (2017) Digitization of herbaria enables novel research. *American Journal of Botany*, **104**, 1281–1284.

Sun, Y. (2023) A Ruiz Orduna, Z Zhang, SJ Feakins, R Jetter, biosynthesis of barley wax β-diketones: a type-III polyketide synthase condensing two fatty acyl units. *Nature Communications*, **14**, 7284.

Sun, Y., **Yao, R.**, **Ji, X.**, **Wu, H.**, **Luna, A.**, **Wang, Z.** ***et al.*** (2020) Characterization of an alkylresorcinol synthase that forms phenolics accumulating in the cuticular wax on various organs of rye (*Secale cereale*). *The Plant Journal*, **102**, 1294–1312.

Talavera, G. & **Castresana, J.** (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, **56**, 564–577.

Tasca, J.A., **Smith, C.R.**, **Burzynski, E.A.**, **Sundberg, B.N.**, **Lagalante, A.F.**, **Livshultz, T.** ***et al.*** (2018) HPLC-MS detection of pyrrolizidine alkaloids and their n-oxides in herbarium specimens dating back to the 1850s. *Applications in Plant Sciences*, **6**, e1143.

Wurtzel, E.T. & **Kutchan, T.M.** (2016) Plant metabolism, the diverse chemistry set of the future. *Science*, **353**, 1232–1236.

Yonekura-Sakakibara, K., **Higashi, Y.** & **Nakabayashi, R.** (2019) The origin and evolution of plant flavonoid metabolism. *Frontiers in Plant Science*, **10**, 943.

Yu, G., **Smith, D.K.**, **Zhu, H.**, **Guan, Y.** & **Lam, T.T.** (2017) Ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, **8**, 28–36.

Zangerl, A.R. & **Berenbaum, M.R.** (2005) Increase in toxicity of an invasive weed after reassociation with its coevolved herbivore. *Proceedings of the National Academy of Sciences*, **102**, 15529–15532.

Zhang, Z., **Wei, W.**, **Zhu, H.**, **Challa, G.S.**, **Bi, C.**, **Trick, H.N.** ***et al.*** (2015) W3 is a new wax locus that is essential for biosynthesis of beta-diketone, development of glaucousness, and reduction of cuticle permeability in common wheat. *PLoS One*, **10**, e0140524.